Int J Life Cycle Assess (2010) 15:321–325
DOI 10.1007/s11367-010-0165-6

COMMENTARY

An unexpected challenge: ionizable compounds in the REACH chemical space

Antonio Franco · Andrea Ferranti · Claus Davidsen · Stefan Trapp

Received: 6 November 2009 / Accepted: 25 January 2010 / Published online: 23 February 2010

Abstract
Purpose About 143,000 industrial chemicals have been pre-registered at the European Chemical Agency for registration according to REACH. The tools, models, and regressions employed for the chemical safety assessment of the registered compounds have limited applicability domains. Thus, it is an important question which fraction of the pre-registered compounds falls into these applicability domains.
Methods A random sample of 1,510 compounds out of the ~117,000 chemicals pre-registered at the European Chemicals Agency and due to registration by 2010 and 2013 was analyzed to investigate the physico-chemical domain of REACH substances. The chemical structure was identified from the CAS number, and the software ACD/Labs was used to calculate dissociation constant(s) (pK_a), octanol–water partition coefficient (log P) and vapor pressure of the neutral molecule.
Results Four hundred ninety-one (33%) of the 1,510 compounds are mostly ionized at pH7 (i.e., acids pK_a<7, bases pK_a>7). Twenty-seven percent of compounds are acids with pK_a<12, 14% bases with pK_a>2, and 8% ampholytes or zwitterionics. Almost half of the ionizable compounds (267 out of 1,510 compounds or 18%) with pK_a between 2 and 12 are even multivalent. There is a high occurrence of hydrophilic chemicals (30% with log P<1), but super-lipophilic chemicals are frequent as well (10% with log P>6). Most chemicals are non- or semi-volatile: the vapor pressure is <1 Pa for 65% and >100 Pa only for 13%.
Conclusions This preliminary characterization of the REACH chemical space helps to identify most urgent gaps of existing in silico tools that are going to be applied in the context of REACH. These data may also be used to select representative sets of test chemicals for the development of new QSARs and models.

Keywords Dissociation constant · Ionization · Lipophilicity · log K_{ow} · pK_a · REACH

A. Franco (✉) · A. Ferranti · C. Davidsen · S. Trapp
Department of Environmental Engineering,
Technical University of Denmark,
Miljøvej, Building 113,
2800 Kgs Lyngby, Denmark
e-mail: anf@env.dtu.dk

1 Introduction

About 143,000 industrial chemicals have been pre-registered at the European Chemical Agency (ECHA) to comply with the EU Regulation for the *R*egistration, *E*valuation, *A*uthorization and restriction of *CH*emicals (REACH; ECHA 2009a). About 117,000 are due to registration by 2010 (>1,000 tons/year; >100 tons/year and very toxic; and >1 ton/year and carcinogenic, mutagenic, or toxic to reproduction) or 2013 (>100 tons/year). The physico-chemical property space of the pre-registered REACH chemicals is yet unknown but is of large interest for risk assessors and regulators concerned with the safety and the fate of those compounds. The tools, models, and regressions employed for the chemical safety assessment of the registered compounds have limited applicability domains. Thus, it is an important question which fraction of the pre-registered compounds falls into these applicability domains. Based on a randomly selected sample of the ~117,000 pre-registered substances due to registration by 2010 and 2013, a physico-chemical characterization of the REACH chemicals space was done and is presented here.

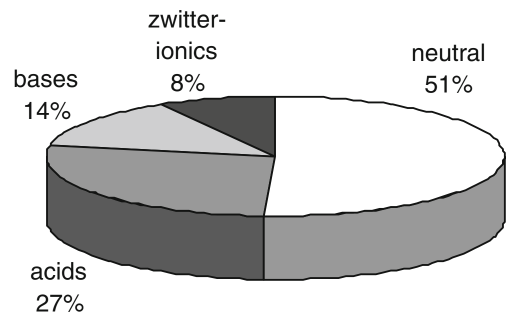

Fig. 1 Percentage of ionics from 1.510 pre-registered REACH chemicals. Acids with $pK_a<12$ and bases with $pK_a>2$ were considered; "zwitterionics" includes amphoteres

2 Methods

A sample of the mentioned ~117,000 pre-registered substances published on the ECHA website (ECHA 2009a) was selected using the random sampling function of Excel®. This initial sample consists of 2,511 entries. The chemical structure was identified from the CAS number using the global suppliers database Chemical Book (http://www.chemicalbook.com, last accessed 16th July 2009) and entered into the software ACD/Labs® (ACD/I-Lab, ver 6.01, Advanced Chemistry Development, Toronto ON, Canada) for the calculation of the dissociation constant(s) (pK_a), the octanol–water partition coefficient of the neutral molecule (log P) and the vapor pressure of the neutral molecule (p_s). About 40% of the substances from the initial sample could not be processed and were thus excluded from the analysis. Major circumstances for exclusion were substances whose structures could not be identified, mixtures that could not be simplified to one single structure and substances that fell outside the applicability domain of ACD (e.g., inorganic chemicals). Salts were considered as dissociated, and the estimate was done for the correspondent neutral organic molecule. The 1,510 processed substances represent 1.3% of all the 117,000 pre-registered substances due to registration by 2010 and 2013.

3 Results

3.1 Ionization

Figure 1 shows the occurrence of ionizable compounds. The usual pH in surface waters is between 6 and 9 (Barndt et al. 1989), but extremes, such as acid lakes and hypertrophic lakes can range from pH4 to 10. Thus, a substance was considered ionizable in the environment if it comprises an acid $pK_a<12$ or a basic $pK_a>2$, i.e., is at least partly ionized in that pH range (pH4 to 10). About one half (49%) of the 1,510 compounds are partly or totally ionized under environmental conditions. The majority of ionizable chemicals are acids (27%) but also bases (14%) and zwitterionics or amphoteres (8%, molecules including both acidic and basic groups) are frequent. About 18% of the total sample comprises multivalent ionics, most of them acids. One third of the total sample (33%, i.e., most of the ionizable) comprises chemicals that are mostly ionized ($pK_{a,acid}<7$ or $pK_{a,base}>7$) at pH7.

Figure 2a reports the distribution of the first acid dissociation constant of acids and zwitterions or amphoteres. The uneven distribution of the acid dissociation constant of the values in the range $-2<pK_a<12$ highlights the occurrence of frequent anionic moieties. There is a high frequency of relatively strong acids with dissociation constants in the range $-2<pK_a<0$, typical for sulfonic and other strong organic acids, and of moderately strong acids with dissociation constants in the range $2<pK_a<5$, typical for carboxylic acids (–COOH). Higher pK_a values are typical for phenols (–OH) and amides (–NH–). There is a relatively low frequency of acids with pK_a in the range $5<pK_a<9$. The basic pK_a (Fig. 2b) is more evenly distributed and no particular pattern can be seen. Very strong bases ($pK_a>11$) are rare.

3.2 Lipophilicity

The octanol–water partition coefficient of the neutral molecule (log P, also known as log K_{OW}), describing the

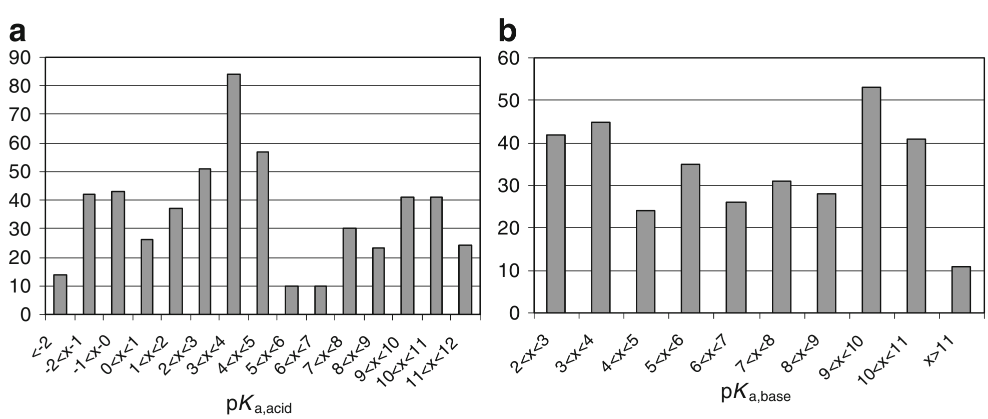

Fig. 2 Distribution of the first acid dissociation constants of **a** acid groups of acids and zwitterionics or amphoteres ($pK_{a,acid}$) and **b** basic groups of bases and zwitterionics ($pK_{a,base}$)

lipophilicity of the analyzed chemicals, varies over more than ten orders of magnitude (Fig. 3). The most frequent log P values range between 0 and 4. There is a high occurrence of hydrophilic chemicals (30% with log $P<1$). Super-lipophilic chemicals are also frequent (10% with log $P>6$). The apparent octanol–water partition coefficient at pH7 (log D), is lower than log P if the chemical ionizes. The substances that are mostly ionized at pH7 (in gray in Fig. 3) are frequently polar in the neutral form but lipophilic ionics occur as well. In particular, 28% of the substances with a log $P>6$, 3% of the total sample analyzed, are mostly ionized at pH7. Long lipophilic structures with a polar ionizable head (e.g., surfactants) fall into this category.

3.3 Volatility

Figure 4 shows the distribution of the vapor pressure of the analyzed chemicals. Most chemicals are non- or semi-volatile: the vapor pressure is <1 Pa for 65% and >100 Pa only for 13%. Again, the actual volatility may be lower due to ionization, because the vapor pressure of ionic molecules is zero.

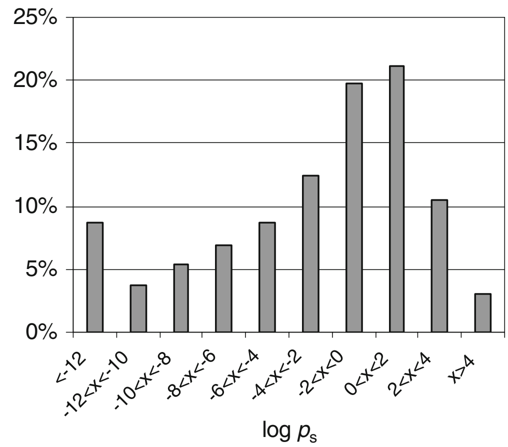

Fig. 4 Distribution of the logarithm of the vapor pressure (log p_s) of the neutral molecule at 25°C (p_s in Pa)

4 Discussion

According to the REACH timeline (ECHA 2009b), compounds with a production volume >100 tons per year need to be registered by 2013, and information on the dissociation constant (pK_a) is required (ECHA 2009c). The registration requires a safety assessment, which is based on testing and model predictions (ECHA 2009b). The high occurrence of ionizable chemicals poses two major challenges to risk assessors: the increased testing requirements and the limited applicability domain of currently used models and regressions.

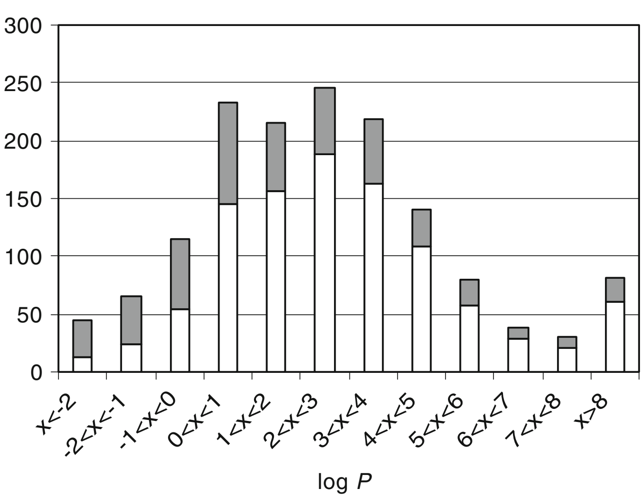

Fig. 3 Distribution of the octanol–water partition coefficient of the neutral molecule (log P) of the 1,510 analyzed pre-registered REACH chemicals. The fraction of chemicals that are mostly ionized at pH7 (p$K_{a,acid}<7$ or p$K_{a,base}>7$) is marked in *gray*

4.1 Data requirements

The guidelines for the chemical safety assessment demand that, for ionizing compounds, "toxicity tests should preferably be carried out at both sides of the pK_a, to fully characterize the possible differences in toxicity" and that "PEC/PNEC comparisons should preferably be made at both sides of the pK_a values, within the environmentally relevant pH range" (ECHA 2009c). This requirement, which holds for a large fraction of the data set, will significantly increase efforts and costs of the assessment. Intelligent testing strategies may reduce the costs and optimize both the effect assessment and the exposure assessment.

The effect assessment includes information on persistence, bioaccumulation potential, and toxicity (PBT assessment). Measurements on one side of the pH may be sufficient for ionizable chemicals if the worse side scenario could be identified in advance.

The bioaccumulation potential of neutral lipophilic molecules is higher than their correspondent ionic form (Fu et al. 2009) due to the higher tendency of the neutral species to cross biological membranes (Trapp and Horobin 2005). This is likely to result in higher bioconcentration factors (BCF) and higher toxicity for the undissociated species as, for example, it was observed for goldfish exposed to chlorophenols (Kishino and Kobayashi 1995, 1996) or fluoxetine in Japanese medaka (Nakamura et al. 2008). Measurements of the BCF and of the toxicity of ionizable chemicals could then be carried out on one side only (i.e., at low pH for acids and at high pH for bases) to account for the worse side scenario.

The persistence of ionizable chemicals is affected by the pH directly, through different uptake of neutral and ionic species, and indirectly, through effects on sorption. These effects may be contrasting. For example, neutral species of organic acids are better taken up by bacteria (Zarfl et al. 2008) but usually exhibit higher sorption (Franco et al. 2009), thus reducing bioavailability. The identification of the worst side scenario for the persistence may therefore be difficult.

Human and environmental exposure assessment is estimated with models covering all potential exposure pathways. There is no clear evidence on which species determines the worst case scenario. Exposure models have been recently refined to account for ionization (Franco and Trapp 2009) and can be run using a standard scenario (e.g., pH 7) for the lower tier assessment and, eventually, running a probabilistic simulation covering the whole range of environmental pH for the higher tier assessment.

4.2 Models applicability

Most regressions and models used in the context of chemicals risk assessment were primarily developed for neutral, lipophilic compounds. For example, the regressions used for the assessment of indirect human exposure are only applicable to neutral, medium lipophilic compounds (Trapp and Schwartz 2000). The predictor variable is the partition coefficient between octanol and water, and the common regression range is only between log P 3.0 to 4.6. This can be contrasted to properties of the random sample analyzed. Out of the set of 1,510 compounds, 268 compounds (18%) have their log P in this range, but only 146 compounds (9.6%) are also essentially non-dissociable (acids with $pK_a<12$ and bases with $pK_a>2$). The application of models and QSARs outside their domain will increase the uncertainty and decrease the reliability of model predictions.

4.3 Limitations of the study

The 1,510 analyzed substances comprise a little more than 1% of all the pre-registered substances. In addition, about 40% of compounds could not be processed or fell outside the range of the property estimation routines of ACD/Labs®. The exclusion of these compounds from the analysis may have biased the sample representativeness. The uncertainty associated with the software ACD/Labs® is acceptable for the purpose of this study. This program is suggested for the estimation of physico-chemical properties by the REACH guidance document (ECHA 2009c). It was evaluated by the EU REACH Implementation Project (ECHA 2009c), and was among the best programs available for the estimation of log P (standard error=0.27) and vapor pressure. It was also tested for its ability to predict pK_a, and was the best of three tested methods with an estimated root mean squared error of 0.41 for acids and 0.82 for bases (Yu et al. 2009).

4.4 Other studies

Manallack (2007) investigated the pK_a distribution of pharmaceuticals using a test set of 582 compounds. The vast majority of drugs (77.5%) had an ionizable group within the relevant range (pK_a 2 to 12), hereof 45.4% had a single base group, 24.4% had a single acid group, 14.8% were ampholytes (11.2% were ampholytes with only one acid and base group), and 10.5% were bivalent bases. The study cites others, which give for the percentage of ionizable drugs numbers between 62.9% and 95%. For the environmental risk assessment of pharmaceuticals, similar tools are applied as for the registration of chemicals (EMA 2006).

5 Conclusions

A large fraction of the pre-registered REACH compounds is ionizable or polar, and thus not inside the applicability domains of estimation methods and QSARs suggested for the chemical safety assessment. This increases the uncertainty of the assessment and the efforts of testing and may even lead to false results. Therefore, QSARs and models developed for hazard and fate assessment of chemicals should include polar and ionizable compounds. A positive aspect is that bioaccumulation and toxicity of ions are lower than that of the corresponding neutral compounds.

Acknowledgments This work received financial support from the European Union 6th Framework Program of Research, Thematic Priority 6 (Global change and ecosystems), contract number GOCE 037017, project OSIRIS. Support for this work was also provided through a Ph.D. grant of the Technical University of Denmark for Antonio Franco.

References

Barndt G, Bohn B, Köhler E (1989) *Biologische und chemische Gütebestimmung von Fließgewässern*. VDG Vereinigung Deutscher Gewässerschutz, Bonn, Germany, 2nd ed., p. 79.

ECHA European Chemical Agency (2009a) ECHA publishes an updated list of pre-registered substances. Press ECHA/PR/09/03. http://echa.europa.eu/doc/press/pr_09_03_list_prereg_substances_20090327.pdf (accessed 14 July 2009)

ECHA European Chemical Agency (2009b) REACH guidance documents. http://guidance.echa.europa.eu/guidance_en.htm (accessed 14 July 2009)

ECHA European Chemical Agency (2009c) Guidance on information requirements and chemical safety assessment. Endpoint specific

guidance Chapter R.7a, p. 173. European Chemicals Agency, Helsinki, Finland

EMA European Medicines Agency, Committee for Medicinical Products for Human Use CHMP (2006) Guideline on the environmental risk assessment of medicinical products for human use. http://www.ema.europa.eu/pdfs/human/swp/444700en.pdf (accessed 15 January 2010)

Franco A, Trapp S (2009) A multimedia activity model for ionizable compounds—validation study with 2, 4-D, aniline and trimethoprim. Environ Toxicol Chem (accepted). doi:10.1002/etc.115

Franco A, Fu W, Trapp S (2009) Soil pH and sorption of ionizable chemicals: effects and modeling advance. Environ Toxicol Chem 28:458–464

Fu W, Franco A, Trapp S (2009) Methods for estimating the bioconcentration factor (BCF) of ionizable organic chemicals. Environ Toxicol Chem 28:1372–1379

Kishino T, Kobayashi K (1995) Relation between toxicity and accumulation of chlorophenols at various pH, and their absorption mechanism in fish. Water Res 29:431–442

Kishino T, Kobayashi K (1996) Acute toxicity and structure-activity relationships of chlorophenols in fish. Water Res 30:387–392

Manallack DT (2007) The pKa distribution of drugs: application to drug discovery. Perspectives in Medicinal Chemistry 1:25–38

Nakamura Y, Yamamoto H, Sekizawa J, Kondo T, Hirai N, Tatarazako N (2008) The effects of pH on fluoxetine in Japanese medaka (*Oryzias latipes*): acute toxicity in fish larvae and bioaccumulation in juvenile fish. Chemosphere 70:865–873

Trapp S, Schwartz S (2000) Proposals to overcome limitations in the EU chemical risk assessment scheme. Chemosphere 41:965–971

Trapp S, Horobin RW (2005) A predictive model for the selective accumulation of chemicals in tumor cells. Eur Biophys J 34:959–966

Yu H, Kühne R, Ebert RU, Schüürmann G (2009) Comparative analysis of different models to predict pK_a values. Proceedings of the 12th EuCheMS International Conference on Chemistry and the Environment, 14–17 June 2009, Stockholm, Sweden

Zarfl C, Matthies M, Klasmeier J (2008) A mechanistical model for the uptake of sulfonamides by bacteria. Chemosphere 70:753–760